HSF-88-01

Preliminary Evaluation of Prototype Lane Control Signs

Ontario

Ministry of Transportation

Research and Development Branch

Technical Report Documentation Page

Preliminary Evaluation of Prototype Lane Control Signs

Author(s):	G.R. Engel, M. Townsend, W. Dougherty, Engel & Townsend Consultants
Number:	HSF-88-01
Date of Publication:	July 1988
Published by:	Research and Development Branch, MTO
Ministry Contact:	C. Blamey (416) 235- 4444
Abstract:	This report describes the results of a field study to evaluate the relative legibilities of four fibre-optic, prototype lane-control signs. The signs could display one of four symbols (three directional arrows and an "X") or one of seven advisory speeds; the characters were about 0.5 m high and ranged in width from 0.3 m to 0.9 m. Forty volunteer observers viewed the signs (which were in-situ on the roof of a Ministry building) from a distance of about 230 m. Under the experimental set-up the observers were asked: to identify the characters displayed on the signs; to describe what they thought the symbols and numbers on the signs meant; and to state which sign best conveyed its message. The analysis of the results showed that two of the four signs provided the best legibility performance; one sign was best for the symbols and the other sign was best overall for the numbers; however the sign best for symbols might have improved its numbers legibility performance if all the numbers had been available on its display. Compared with standards for constructing legible printed signs for use in other environments, the signs that were tested in this trial used character sizes that were significantly smaller than would be expected for reliable legibility performance. The designs of these signs should be tested in a laboratory where the several variables that affect legibility (such as font design, character-size, luminance, colour, etc.) can be properly controlled and manipulated.
Comments:	MTO client: Freeway Traffic Management Systems
Key Words:	freeway, signs, fibre-optic, lane control, evaluation, human factors
Distribution:	Internal and on request.
Copyright Status:	Crown copyright © 1988 Ministry of Transportation

HSF-88-01

Preliminary Evaluation of Prototype Lane Control Signs

G.R. Engel

M. Townsend

W. Dougherty

Engel & Townsend Consultants

Published by
The Research and Development Branch
Ontario Ministry of Transportation
Hon. Ed Fulton, Minister
D.G. Hobbs, Deputy Minister

Published without prejudice as to the application of the findings.
Crown copyright reserved; however, this document may be reproduced for non-commercial purposes with attribution to the Ministry.

For additional copies, contact:
The Editor, Technical Publications
Room 320, Central Building
1201 Wilson Avenue
Downsview, Ontario
Canada M3M 1J8

Telephone: (416) 235-3480
CNCP Dialcom Electronic Mail: 23:MIN006

July 1988

Table of Contents

List of Tables and Figures iii
Acknowledgements iv

1/ EXECUTIVE SUMMARY v
1.1/ The Prototype Signs v
1.2/ Evaluation Design v
1.3/ Conclusions v

2/ INTRODUCTION 1

3/ EVALUATION CONDITIONS 1
3.1/ The Lane Control Signs 2
3.2/ The Site 8
3.3/ The Viewing Apparatus 11
3.4/ Subjects 11
3.5/ Weather Conditions 12

4/ EVALUATION DESIGN 12
4.1/ Legibility 12
4.1.1/ Symbols 14
4.1.2/ Numbers 16
4.2/ Comprehension 17
4.3/ Preference 17

5/ RESULTS 18
5.1/ Legibility 18
5.1.1/ Symbols Legibility 18
5.1.2/ Numbers Legibility 27
5.2/ Comprehension 34
5.3/ Preference 36
5.4/ Characteristics of Legible Signs 39

6/ DISCUSSION 42
6.1/ Conclusions 50

7/ REFERENCES 52

List of Tables and Figures

TABLES

1/ Dimensions of Signs and Characters 3
2/ Confusion Matrices for the Symbols 19
3/ Comparisons of Manufacturer's Signs: Accuracy, Information and Chi Squares for the Symbols 21
4/ Confusion Matrices for the Symbols: Corrected for Response Bias 25
5/ Confusion Matrices for the Numbers 28
6/ Comparisons of Manufacturer's Signs: Accuracy, Information and Chi Squares for the Numbers 29
7/ Confusion Matrices for the Numbers Corrected for Response Bias 31
8/ Sign Displays Ordered According to Accuracy 38

FIGURES

1/ Illustrations of Characters used on the Sylvia Sign 4
2/ Illustrations of Characters used on the Futurit Sign 5
3/ Illustrations of Characters used on the Yvan Paque Sign 6
4/ Illustrations of Characters used on the Dambach Sign 7
5/ Photograph of Sign Display and Backboards 9
6/ Location of the Prototype Signs and the Viewing Trailer 10
7/ Original and Corrected Accuracies for the Dambach and Sylvia Symbols 15
8/ Accuracies for Sylvia, Yvan Paque and Dambach Numbers 33
9/ Number and Symbol Preferences 37

ACKNOWLEDGEMENTS

We thank Chris Blamey, of the Highway Innovations and Strategic Research Office, who was the Project Manager for this study, for his intelligence and efficiency in helping us to perform our work.

We thank Tuan Tien who along with Wally Dougherty acted as field researchers for the project.

We thank Mike Lau, Project Engineer, Traffic Management and Engineering Office, for his cooperation throughout the project.

And finally, we thank the members of the public who acted as observers during the trials, for their cooperation and perseverance throughout the three hours it took to run a trial.

1.0 EXECUTIVE SUMMARY

This report describes the results of a field study to evaluate the relative legibilities of four prototype lane control signs. Lane control signs are intended to be part of a freeway traffic management system for Highway 401 in the Metropolitan Toronto area.

1.1 The Prototype Signs

The trade names of the four manufacturer's signs were Sylvia, Futurit, Yvan Paque and Dambach. The prototype fibre optic signs displayed four symbols and seven numbers. The symbols were; a green arrow pointing down, an amber arrow facing left, an amber arrow facing right, and a red 'X'. The numbers were; 30, 50, 60, 70, 80, 90, displayed in white, yellow, or yellow-amber, and a green 100. The signs could be displayed in a flashing or steady mode; except for the Dambach sign, which did not have a flashing mode, nor did it display a 60 or 70. The characters displayed on the signs were about 0.5 metres high and ranged from 0.3 to 0.9 metres wide.

1.2 Evaluation Design

For ongoing environmental tests, the signs had already been mounted, side by side, on the roof of a building in the Downsview complex. Forty volunteer observers viewed the signs from a distance of 233 metres. The observers were asked to identify the characters displayed on the signs while viewing the signs through a filter that degraded the scene. This task provided data on the relative legibilities of the signs. Observers were also asked to describe what they thought the symbols and numbers on the signs meant. This task provided comprehension data. And finally a measure of viewer preference was obtained by asking the observers which sign best conveyed its message.

1.3 Conclusions

The analysis of the results yielded the following conclusions:

1) The Dambach sign gave the optimum legibility performance for the symbols characters.

2) The Yvan Paque sign gave the optimum legibility performance for the numbers. The Dambach sign would also be a possible choice for the number displays provided that it could be shown that the 60 and 70 displays which were not provided by Dambach for the trial, yielded legibility performance equal to that of the other Dambach displays.

3) For a choice of one manufacturer to provide both the symbol and the numeric displays, there were two candidates neither of which would be a clear choice on the basis of this trial. Yvan Paque would be one choice provided

that Yvan Paque could implement the solid geometric form type of symbol display. Dambach would be the other choice provided that Dambach could implement satisfactory 60 and 70 sign displays.

4) The trial results suggested that the best font style for the number signs is a line stroke style constructed from closely spaced single lines of light elements. For the symbol signs, dense arrays of light elements forming solid geometric shapes appear to provide the best legibility.

5) The trial showed that the flashing displays significantly decrease legibility; and that the flashing mode does not convey any additional information not available in the steady mode.

6) Although it was not possible to systematically measure the effects of sign colour in the trial, it was evident that colour did not convey any extra information; and to the degree that other research can be generalized to the message signs, colour probably decreased the legibilities of the signs.

7) Observers interpretations of the meanings of different signs were not consistent; and were often quite incorrect. This suggests that drivers would have to be educated about the meanings of the signs, and educated about driving behaviors that would be appropriate in response to the sign messages.

8) Compared with standards for constructing legible printed signs for use in other environments, the signs tested in this trial use character sizes that are about half the size that would be required for reliable legibility performance. In fact, at 300 metres, the number signs are equivalent to reading relatively small print on a printed page. Considering the environment in which the signs are to be used, there is little reason to be optimistic about their legibility characteristics. The obvious solution to this problem would be to increase the size of the signs, and to make them as bright as possible. If there are limits to the size and brightness of the signs, careful attention should be given to the font designs used for the signs.

9) The designs of the signs should be tested in a laboratory setting where the several variables that affect legibility can be properly controlled and manipulated. Laboratory testing would be critical to optimising such things as font design, colour, and size-luminance relationships.

Preliminary Evaluation of
Prototype Lane Control Signs

2.0 INTRODUCTION

This report describes the results of a study to evaluate four prototype lane control signs. Lane control signs are an integral part of the development of a traffic management system for Highway 401 in the Metropolitan Toronto Area. Changeable message signs mounted on gantries over the freeway will convey information about general traffic conditions and problems to motorists. The messages contained on changeable message signs will consist of short strings of text. Lane control signs on the other hand, are signs that convey information about a specific lane of traffic on the freeway. This information is in the form of simple numerical displays and symbols designed to convey speed and lane use information.

Four manufacturers submitted prototypes of lane control signs. The signs were mounted, side by side, on the roof of a building in the Downsview complex. The preliminary study was designed to evaluate these signs for legibility, comprehension and viewer preference.

The report will begin with a brief description of the signs and the setting in which the evaluation was carried out. Following this, we will describe the design of the evaluation. Next the results of the data analyses will be presented. Finally we will present our conclusions and recommendations.

3.0 EVALUATION CONDITIONS

In this section we will briefly describe the prototype lane control signs and how they were displayed. Then we will describe the general set up of the evaluation, including the apparatus that was used to view the signs. A brief

description of the volunteer observers and the weather conditions under which the evaluation was run will complete this section.

3.1 The Lane Control Signs

Four manufacturer's provided prototype lane control signs. The trade names of the signs were; Sylvia, Futurit, Yvan Paque and Dambach. Each sign contained fibre optic light bundles that could be controlled by a micro-processor so that various numbers and symbols could be displayed on command. Filters in the signs allowed the signs to display the numbers or symbols in colour; either red, amber or green. The signs could also be dimmed and, with the exception of the Dambach sign, were also capable of being programmed to flash off and on. At the time of writing this report, a separate study was being conducted to determine each sign's physical, operational and performance characteristics. Therefore the luminance levels, contrast ratios, flash rates, and other physical properties of the signs were not available for inclusion in this report or for use in the analysis of the data.

The signs were about 0.8 metres high and about 1.2 metres wide. The fibre optic characters that were displayed on the signs ranged from about 0.4 to 0.6 metres high and anywhere from 0.3 to 0.9 metres wide. The exact dimensions of each of the signs and the size of the various characters contained on the signs are listed in Table 1. The sizes of the fibre optic characters varied considerably within each manufacturer's sign, and between manufacturer's signs. Although the Ministry provided height/width specifications for the display characters, it appears the manufacturers did not closely follow these specifications.

Figures 1 through 4 illustrate the character sets for each of the manufacturer's signs. The drawings were made by projecting a photographic slide of each character onto a piece of paper and marking each light source with a dot. Therefore, the number of dots illustrated shows the exact arrangement of fibre optic light sources for each character. The colour of each character is noted below each tracing. The tracings also illustrate the placement of the character within the sign. For example, some manufacturers placed the arrows in the right and left corners of their signs. The major

Table 1: Dimensions of Signs and Characters

Sign	Size(mm)	Character	Character Size(mm)
SYLVIA	1140 x 840	100	419 x 806
		90	530 x 726
		80	530 x 726
		70	530 x 726
		60	530 x 726
		50	530 x 726
		30	530 x 726
		X	530 x 530
		Left Arrow	447 x 447
		Right Arrow	447 x 447
		Down Arrow	475 x 475
FUTURIT	1145 x 845	100	450 x 810
		90	430 x 770
		80	450 x 760
		70	440 x 750
		60	440 x 740
		50	450 x 730
		30	430 x 710
		X	560 x 570
		Left Arrow	580 x 580
		Right Arrow	580 x 580
		Down Arrow	560 x 580
YVAN PAQUE	1160 x 860	100	500 x 950
		90	unknown
		80	500 x 810
		70	430 x 690
		60	400 x 650
		50	500 x 720
		30	500 x 740
		X	480 x 350
		Left Arrow	510 x 490
		Right Arrow	490 x 490
		Down Arrow	390 x 310
DAMBACH	1140 X 840	100	500 x 840
		90	510 x 770
		80	500 x 790
		50	500 x 860
		30	500 x 730
		X	430 x 430
		Left Arrow	400 x 380
		Right Arrow	400 x 380
		Down Arrow	480 x 480

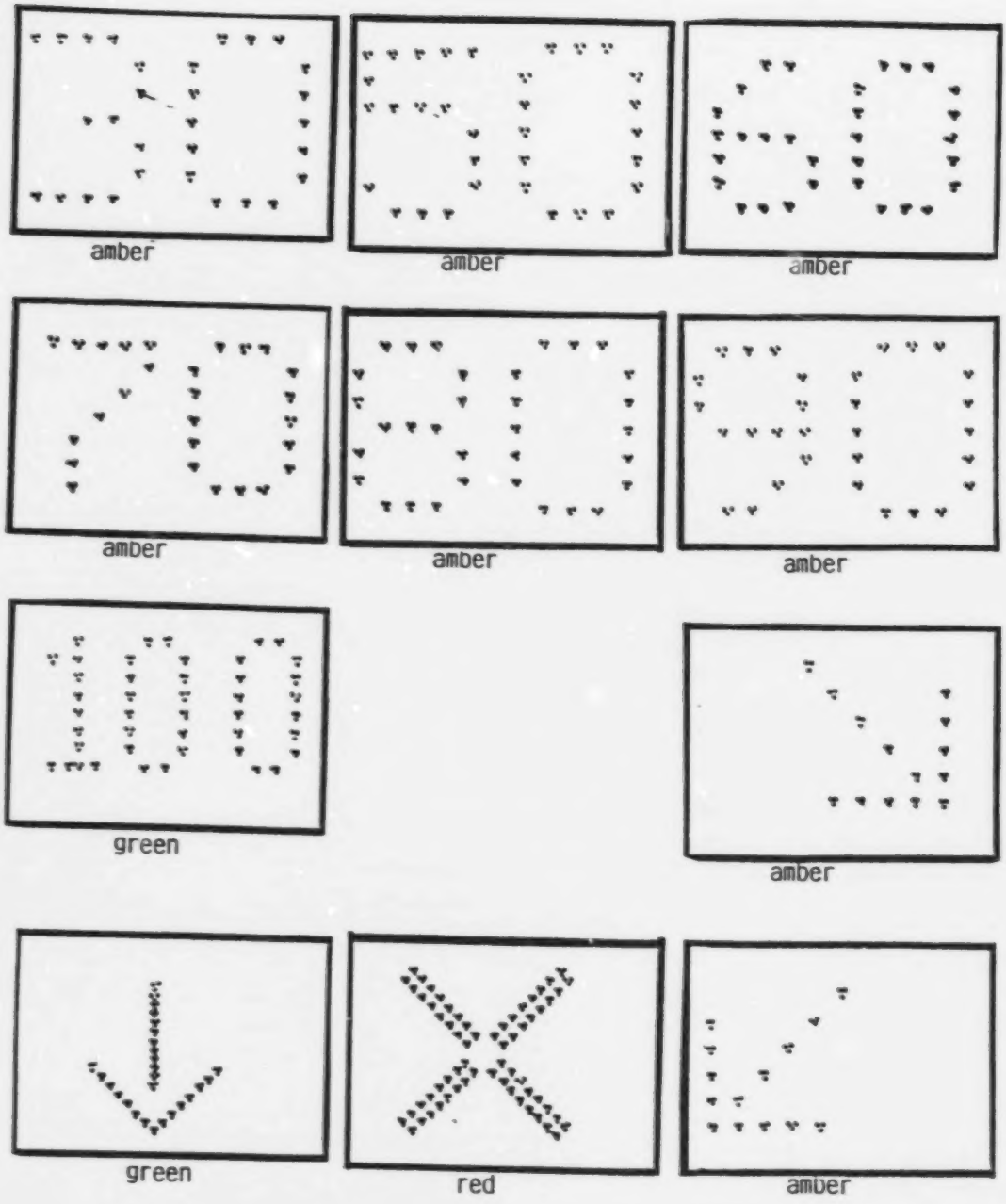

Figure 1: Illustrations of Characters used on the Sylvia Sign

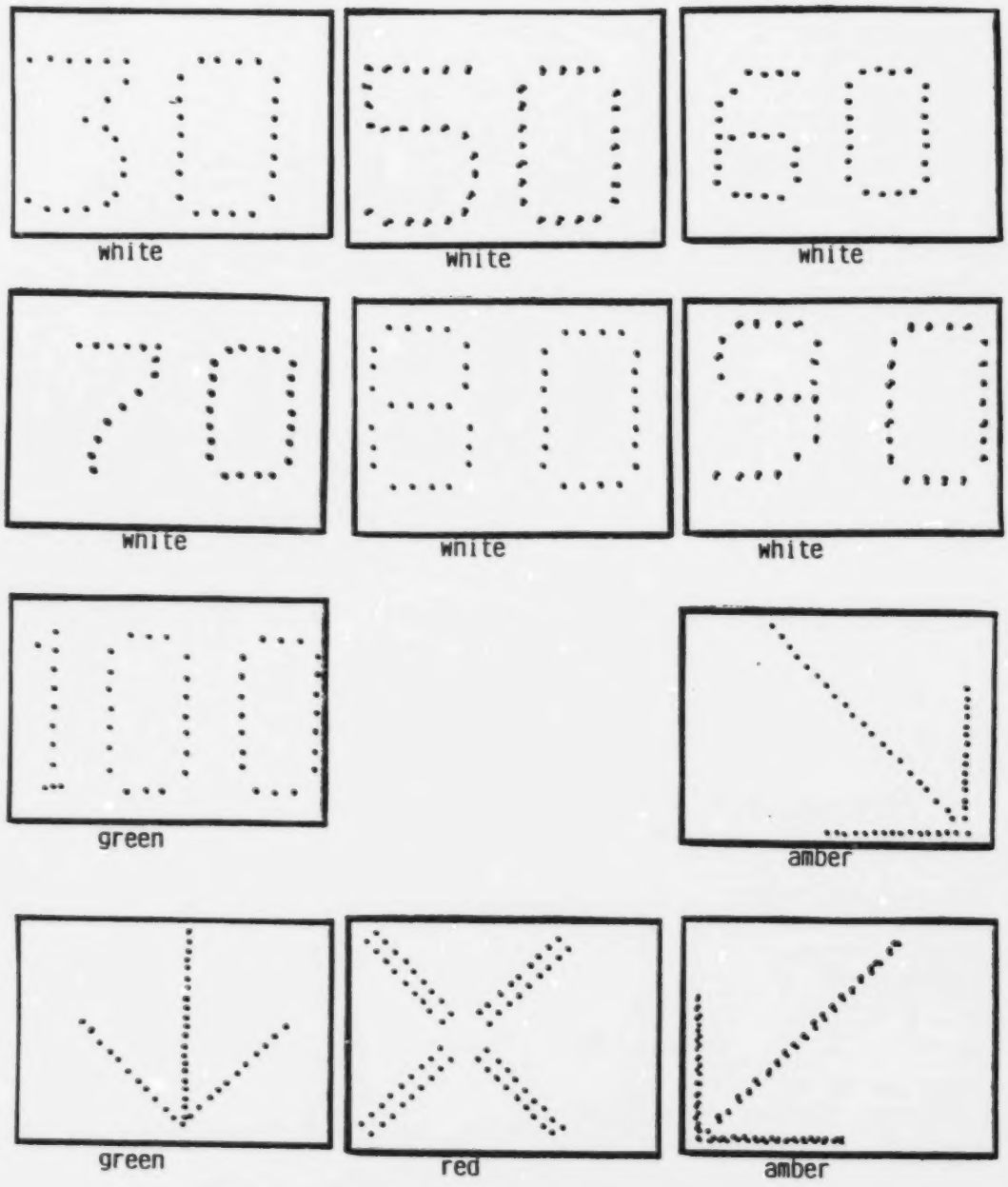

Figure 2: Illustrations of Characters used on the Futurit Sign

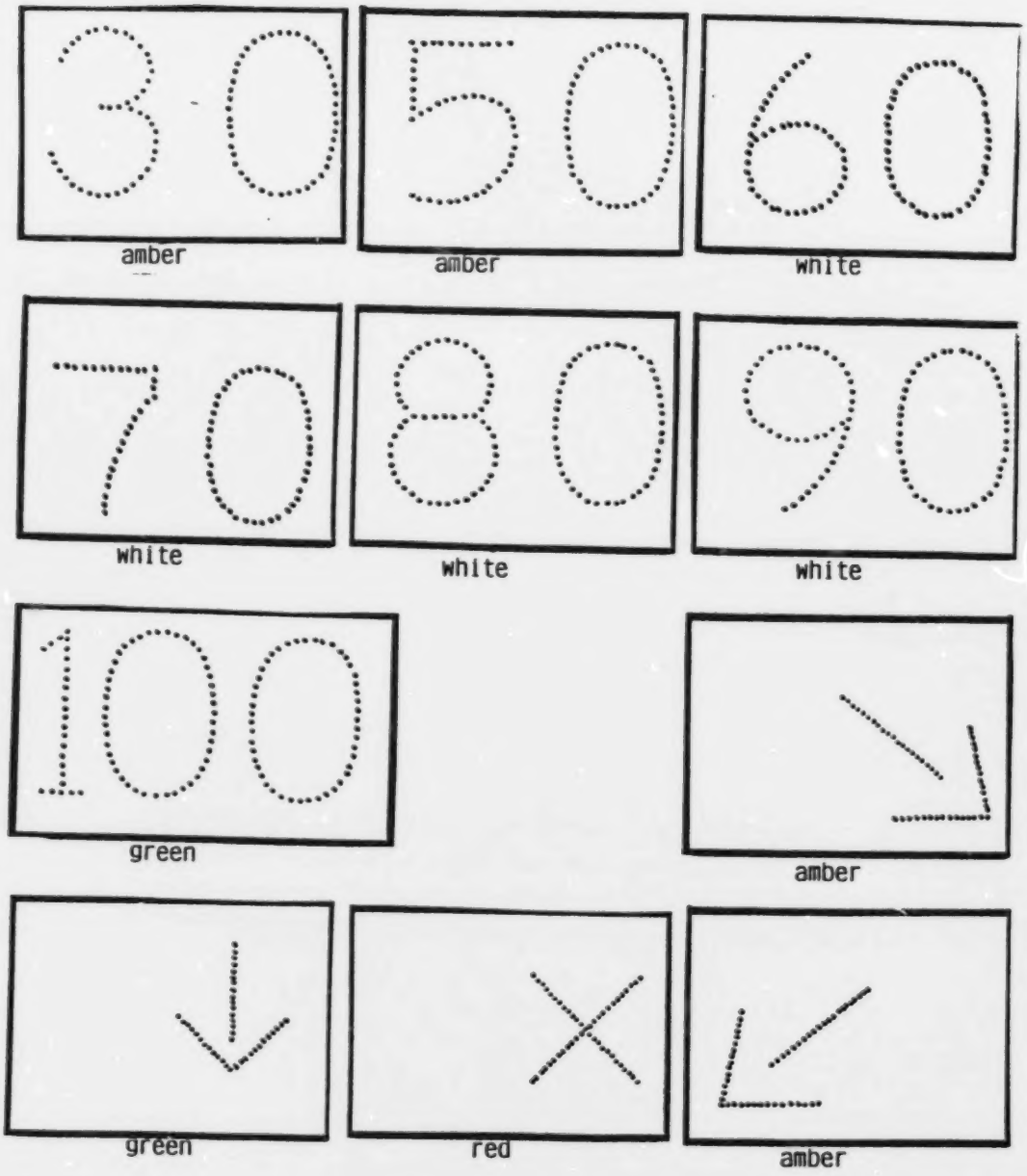

Figure 3: Illustrations of Characters used on the Yvan Paque Sign

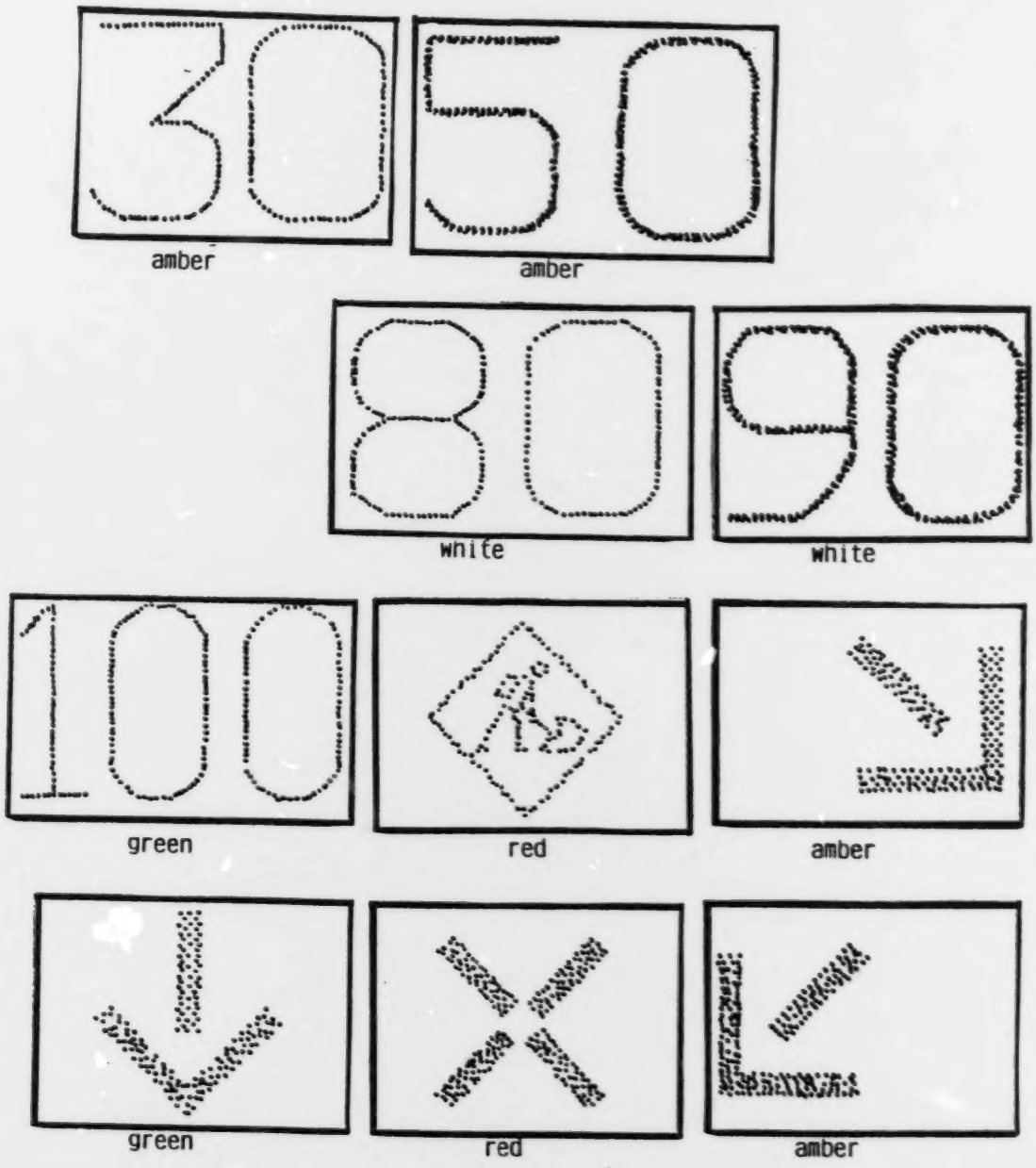

Figure 4: Illustrations of Characters used on the Dambach Sign

differences among the character sets was in the thickness of the character lines and the placement of the characters relative to the borders of the signs.

3.2 The Site

The signs were mounted, side by side, on the roof of a building in the Downsview complex. From the location where the signs were to be viewed, they appeared to be below the horizon, and so a row of black panels was mounted behind the signs to minimize any distractions from the city beyond. Figure 5 shows how the signs were displayed on the roof of the Carpenter and Paint Shop Building. The backboard was 8 metres long and 1.9 metres high.

The controller mechanisms for each sign were in a room at the back of the building. A research assistant sat in this room to operate a computer that controlled the sign displays during the evaluation test sessions.

Another research assistant and a volunteer observer, sat in a box trailer that was parked 233 metres away from the signs on the roof. Although the functional requirements for the signs specified that they were to be visible up to 300 metres, 233 metres was the longest unobstructed viewing position that could be found on the site. Figure 6 is a map of the Downsview Complex, and the location of the signs and the trailer are marked.

There was also a construction site near the trailer. Movement of heavy vehicles into and out of this site occasionally obstructed the view. When this occurred, the sequence of viewing events was interrupted and had to be re-started or re-scheduled. On dry days, breezes stirred up dust from the construction site, and this obstructed the view for a few minutes as well. These events were random and unpredictable, and we do not believe that they affected the outcome of the evaluation in any systematic way.

The research assistants communicated with each other via a walkie talkie.

Figure 5: Photograph of Sign Display and Backboards

Ministry of Transportation
Ontario

M.T.C. DOWNSVIEW COMPLEX

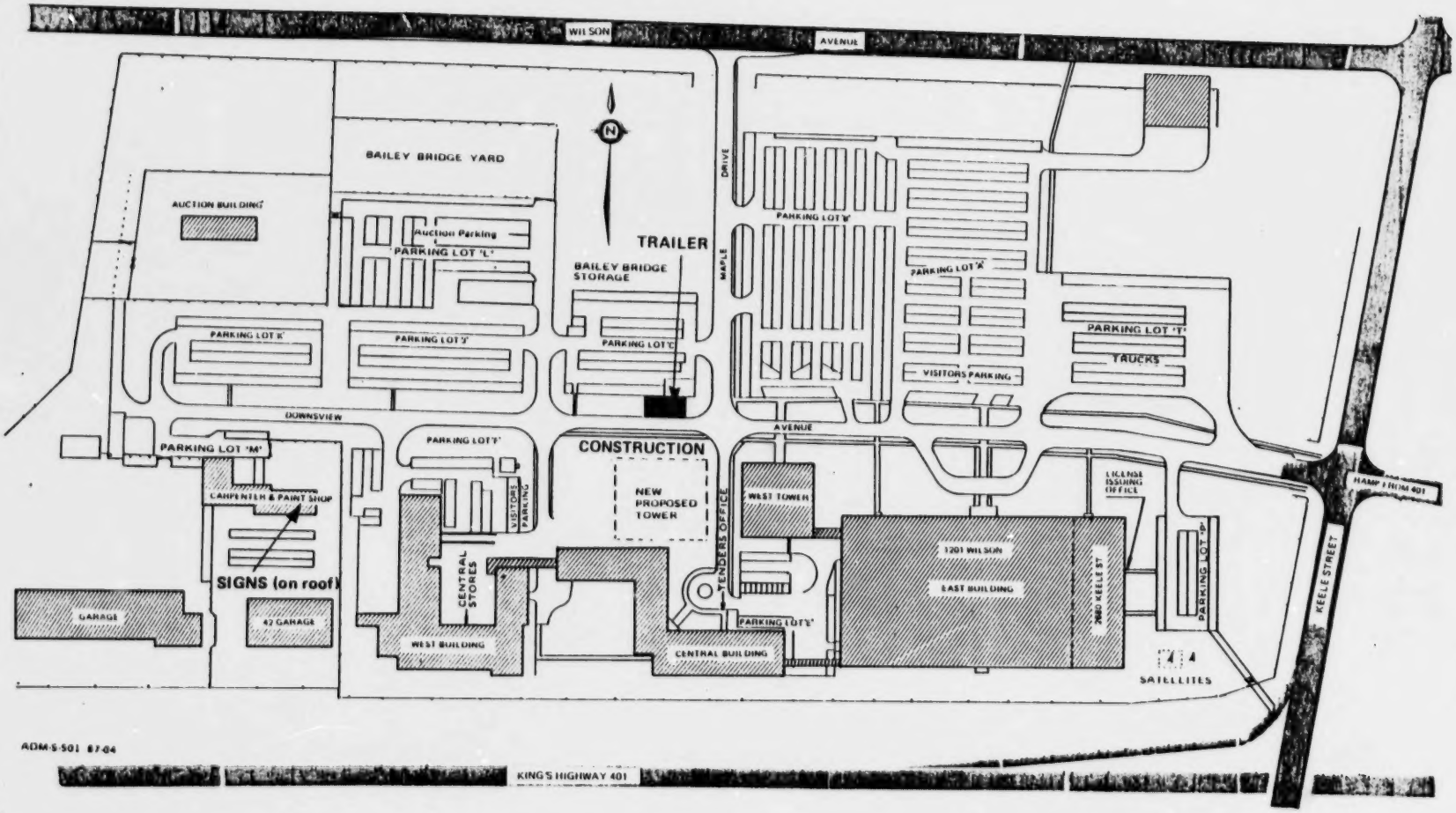

Figure 6/ Location of the Prototype Signs and the Viewing Trailer

3.3 The Viewing Apparatus

The viewing apparatus consisted of a neutral density wedge filter (1) that was 2 cm wide and 15 cm long, going from density 0 to density 4. A density of 4 reduces the light transmitted through the filter by a factor of one ten thousandth. The wedge was mounted horizontally on a microscope stage, which in turn was mounted on an adjustable camera tripod. A thin piece of bristol board with a small viewing aperture about 6mm in diameter was mounted on the tripod as well. By turning an adjusting knob on the side of the microscope stage, the density wedge filter could be moved back and forth in front of the aperture to control the amount of light coming through the aperture. The observer sat in a chair behind the viewing apparatus.

3.4 Subjects

Fifty volunteers took part in the study. Five people participated in a pilot study, 40 people participated in the evaluation study, and remaining five took part in a post-evaluation study.

Volunteers for the evaluation came from two major sources. Some were friends and acquaintances of the researchers. The remainder were recruited by means of a notice placed in the licence renewal office at the Downsview complex. All volunteers were scheduled for a morning or afternoon session,

---

1 A neutral density filter reduces absolute brightness, that is, brightness relative to black, by an amount corresponding to the density of the filter. In the case of gray scale brightnesses, all brightness levels are reduced in proportion to the density of the filter, while contrast ratios among levels of gray remain constant. Brightnesses of different colours are affected equally relative to the density of the filter. Thus a neutral density filter does not affect the colour composition of the scene. In a neutral density wedge filter the density varies continuously along the length of the filter.

and were called the day before a session to ensure that they would keep their appointments. They were paid $20.00 per hour for their participation. The evaluation procedure took approximately three hours to complete.

The average age of the volunteers was 33.9 years. The standard deviation was 13.1 years. The youngest volunteer was 18 and the oldest was 68. The average age for drivers in Ontario is about 41-42 years old, so our volunteers are somewhat younger than average. Sixty percent of our volunteer sample were males and 40% were females. The percent of male and female drivers in Ontario is 56 and 44 respectively.

3.5 Weather Conditions

The evaluation trials began at the end of August and were completed in November. Weather conditions were recorded throughout the evaluation. Overall, 18% of the evaluation trials were run on sunny days, 48% were run on days where there was solid cloud cover or rain, and the remaining 34% were run on days where the weather was unsettled and varied between sun and cloud. Thus, the trials were run in weather conditions that would simulate driving conditions in Ontario at this time of the year.

4.0 EVALUATION DESIGN

This section contains details of the methods used to measure legibility, comprehension and viewer preference.

4.1 Legibility

Legibility is a measure of the clarity of text; and is usually measured by the accuracy with which text or symbols can be read under given viewing conditions. It is a function of character size, luminance, contrast ratio, colour, font type, viewing distance as well as other factors known to affect visibility. Legibility can be measured in one of three ways; the threshold

distance at which material can be read accurately, the threshold exposure time at which material can be read accurately, or the threshold brightness at which material can be read accurately.

The specifications for the design of the prototype lane control signs were that they were to be legible up to 300 metres. One way to determine which manufacturer's character displays were the most legible would be to see which characters could be read at the greatest distances. Because of the physical limitations of the site, it was not possible to measure legibility by moving observers closer to or further from the signs. Threshold studies also require multiple observations to determine the legibility threshold, and the time required to either drive, walk, or cycle observers back and forth would be exorbitant.

The second method of measuring legibility in a field experiment like this one, might be to have observers view the signs through a variable speed shutter, and measure the exposure times needed to accurately read the characters. This method was explored in a pilot test. Unfortunately, the timer mechanism and the shutter did not operate reliably under these field conditions. Another problem with using a shutter is that it is difficult to give the observer a fixation or focus point before exposing the sign. This could lead to missed observations because the observer's eyes were not properly focused for observing at the viewing distance required.

The practical problems of using distance or exposure time for measuring legibility left brightness as varied by neutral density filters as the only way to vary viewing conditions. A pilot study was conducted to examine the feasibility of using this method. The apparatus previously described was used. The results of the pilot work gave an indication of how long a series of trials would take, how effectively the two researchers could communicate with each other to run the trials in the most efficient manner, how fatiguing the procedure would be for the observers, how long the signs should be viewed and what range of densities would generate an acceptable level of correct responses and errors. The observers were asked to give a response to every exposure of a character, including guessing if they were not sure which character had been presented. This is called a forced choice method. This method allowed us to examine visual confusions among the sign characters.

Because of individual differences between observers there was no single setting of the neutral density filter that would produce roughly the same recognition accuracy for every observer. Consequently, it was necessary to calibrate a setting for each observer.

The observer sat in the trailer with one of the researchers. The height of the viewing apparatus was adjusted for comfort. A sign was exposed for eight seconds and the observer was asked to identify the character on the sign. The neutral density wedge was adjusted so that the overall accuracy of identifying characters was about 50 to 60% on the Sylvia and Yvan Paque signs.

The Futurit sign appeared to be quite dull, while the Dambach sign appeared to be quite bright. If the filter was adjusted to yield 50 to 60% accuracy for the Futurit sign there would be 100% accuracy for the bright Dambach sign. If the filter was adjusted to yield 50 to 60% accuracy for the Dambach sign, accuracy would be close to zero for the dull Futurit sign. So the two signs in the middle range, the Sylvia and Yvan Paque signs, were used to set the filter density for each subject. Once the density level was determined based on this method, the density was not changed for the observer for the rest of the trials.

Because of small variations in visual acuity, the filter density level varied slightly between observers, but on average the density was about 3.5. Looking through the filter at this density would be like viewing a scene just after sunset.

4.1.1 Symbols

Each manufacturer's sign contained four lane control symbols; a green arrow pointing down, an amber arrow facing left, an amber arrow facing right, and a red X.

A single presentation consisted of an 8 second exposure of a symbol followed by 3 seconds of darkness. The exposure would be either in the steady or flashing mode. The 8 second presentation and 3 seconds between

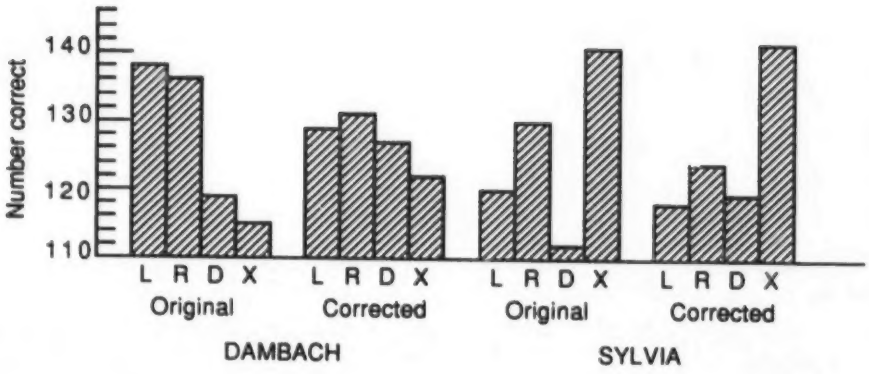

Figure 7/ Original and Corrected Accuracies for the Dambach and Sylvia Symbols

presentations was determined to be a comfortable pace during the pilot testing. In a driving situation, a vehicle moving at 100 km per hour would travel about 220 metres in 8 seconds.

The symbols were presented in blocks of 16 random order presentations (4 symbols, presented 4 times each). A block would consist of only one manufacturer's sign, presented in either the steady, or flashing mode; except for the Dambach sign, which did not have a flashing mode. The observer was given a short break after each block of trials. The observer's response to each presentation was recorded, and as previously noted, the observer was required to give an answer to every presentation, even when he or she was not positive about the identification of the symbol. In total it took about 40 minutes to complete the trials for the symbol presentations.

4.1.2 Numbers

The Sylvia, Futurit and Yvan Paque signs contained the digit sets; 30, 50, 60, 70, 80, 90, 100. The 100 was always green and the other digits were either white, amber or yellow-amber, depending on the manufacturer's design. The Dambach sign did not have the 60 or 70, and it did not have a flashing mode.

The presentation trials were 8 seconds long with a 3 second inter-trial interval. The digits were presented in blocks of trials, similar to the symbol presentation blocks. The 7 digit sets were presented 4 times each in a random order. The Dambach presentations only had 5 digit combinations presented 4 times each during a block. Only one manufacturer's sign was presented at a time, in either the flashing or steady mode, except the Dambach, which again had no flashing mode.

The observers responses were recorded after each presentation, and the whole set of digit presentations took about an hour to complete.

4.2 Comprehension

Legibility and comprehension are independent characteristics of text. For example, a set of characters may be completely legible, but if the characters are in an unknown language, the reader could not possibly comprehend the message.

Lane control signs must be designed so that they clearly indicate what the driver must do. To measure comprehension, we simply asked the observers to tell us what they thought the signs meant. They were told that the signs would be on gantries above the lanes on the Highway 401.

All four manufacturers signs were presented simultaneously for about 30 seconds. They were presented in the steady mode at the daylight illumination level. No filter was used. All four symbols, and the digits, 30 and 100 were presented. The observer's responses were recorded. The Dambach symbol which was designed to represent road construction was also included in the presentations.

4.3 Preference

If all of the manufacturer's signs were equally legible, and the characters were equally effective at conveying the intended message, the only thing that might determine a choice of design is an opinion about which sign conveys the message best.

A measure of viewer preference was obtained by asking the observers to decide which character, in their opinion best conveyed the message. All of the manufacturer's signs were displayed simultaneously for about 20 seconds. The observers gave their preferences by ranking the displays from best to poorest. All of the digits and symbols were ranked in this manner.

To familiarize the observers with the various sign characters, the preference and comprehension ratings were completed before the legibility data were collected.

5.0 RESULTS

The results will be presented in the following order; the legibility data, the comprehension data, and then the preference data.

5.1 Legibility

The legibility study provided data about which manufacturer's sign was more legible than the others, which symbols and numbers were the most legible, and which characters were most likely to be confused with each other. The results for the symbols will be presented first, followed by the results for the numbers.

5.1.1 Symbols Legibility

The data that were collected in the legibility study were tabulated in confusion matrices showing the number of times that observers said they saw particular symbols as a function of the numbers of times that the different symbols were presented. These matrices formed the basic data for all of the analyses that were performed.

Table 2 shows the confusion matrices for the symbols. Except for the Dambach sign, there are two matrices for each manufacturer's sign; one for the presentations made in the steady mode and one for the presentations made in the flashing mode. The Dambach sign had no flashing mode. The number in each cell of a matrix is the number of times observers identified the symbol named in the row heading, when the symbol named in the column heading was actually presented. The column totals are how many times each symbol was presented. Each symbol was presented 160 times, except in the case of the Yvan Paque symbols. Because of an equipment failure during one of the trials, the Yvan Paque symbols were presented 156 times each instead of 160 times. The row totals in the matrices are how many times each symbol was identified by the observer, irrespective of what symbol was presented.

The numbers in the diagonal cells from the upper left corner to the lower

Table 2: Confusion Matrices for the Symbols

STEADY MODE

Symbol Displayed

SYLVIA

	Left	Right	Down	X	SUM
Responses					
Left	119	13	20	11	163
Right	28	130	21	4	183
Down	6	8	112	3	129
X	7	9	7	142	165
SUM	160	160	160	160	640

FUTURIT

	Left	Right	Down	X	SUM
Responses					
Left	124	17	21	38	200
Right	18	118	20	28	184
Down	4	14	103	31	152
X	14	11	16	63	104
SUM	160	160	160	160	640

YVAN PAQUE

	Left	Right	Down	X	SUM
Responses					
Left	77	39	31	34	181
Right	27	56	30	50	163
Down	23	39	75	45	182
X	29	22	20	27	98
SUM	156	156	156	156	624

DAMBACH

	Left	Right	Down	X	SUM
Responses					
Left	138	15	19	25	197
Right	15	136	14	14	179
Down	3	2	119	6	130
X	4	7	8	115	134
SUM	160	160	160	160	640

FLASHING MODE

Symbol Displayed

SYLVIA

	Left	Right	Down	X	SUM
Responses					
Left	109	14	22	14	159
Right	20	119	23	13	175
Down	19	6	102	6	133
X	12	21	13	127	173
SUM	160	160	160	160	640

FUTURIT

	Left	Right	Down	X	SUM
Responses					
Left	109	28	23	31	191
Right	34	83	35	40	192
Down	12	27	77	33	149
X	5	22	25	56	108
SUM	160	160	160	160	640

YVAN PAQUE

	Left	Right	Down	X	SUM
Responses					
Left	66	41	30	43	180
Right	41	45	54	46	186
Down	24	42	51	37	154
X	25	28	21	30	104
SUM	156	156	156	156	624

SUMS ACROSS ALL SIGNS

	Left	Right	Down	X	SUM
Responses					
Left	742	167	166	196	1271
Right	183	687	197	195	1262
Down	91	138	639	161	1029
X	96	120	110	560	886
SUM	1112	1112	1112	1112	4448

right corner of each matrix show the numbers of correct identifications. For example, values in the diagonal cells of the Sylvia matrix are considerably higher than values in the off-diagonal cells, showing that observers were quite accurate on the Sylvia sign. However, for the Yvan Paque sign, the off-diagonal cell values are only slightly lower than the diagonal cell values, showing that the observers were not very accurate on the Yvan Paque sign.

The matrix at the bottom right corner of the table is a confusion matrix aggregated across all of the signs and presentation modes. The data were collapsed across presentation modes because the confusion matrix for the flashing mode had essentially the same pattern of responses as the confusion matrix for the steady mode. The diagonal cells show that overall accuracy ranged from about 50% to about 70%, with an average of about 60%. The row totals show a substantial response bias with a tendency to identify left and right arrow symbols about 30% more often than the down arrow or X symbol, even though all symbols were presented equally often.

Table 3 gives the results of the analyses that were done to compare the legibilities of the symbol characters across manufacturers' signs. The first four rows of the table show the percent of correct identifications for each symbol for each manufacturer's sign. The numbers in parentheses give accuracies for the flashing mode. The numbers without parentheses are for the steady mode.

The last column of accuracy results for each symbol shows accuracy for each symbol averaged across manufacturers. The two rows of results below the results for the individual symbols show the average accuracy for each manufacturer's symbols, and the corresponding standard deviations of these average accuracies.

All of the accuracy values are corrected for guessing so that later on, the identification accuracies for the symbols can be compared with the identification accuracies for the numeric displays. An observer had a one in four chance of guessing a symbol correctly; but only from a one in seven to one in five chance of correctly guessing one of the numeric characters.

The symbols on the Dambach and Sylvia signs showed the highest overall

Table 3: Comparisons of Manufacturer's Signs: Accuracy, Information and Chi Squares for the Symbols
(Numbers in Brackets are for Flashing Characters)

	MANUFACTURER				
Symbols	Dambach	Sylvia	Futurit	Yvan Paque	Average
Left Arrow	61 (na)*	49 (43)	53 (43)	24 (17)	47 (30)
Right Arrow	60 (na)	56 (49)	49 (27)	11 (4)	44 (36)
Down Arrow	49 (na)	45 (39)	39 (23)	34 (8)	42 (23)
Red X	47 (na)	64 (54)	14 (10)	-8 (-6)	29 (27)
Average Accuracy	54 (na)	53 (46)	39 (26)	15 (6)	40 (27)
Standard Deviation	6 (na)	7 (6)	15 (12)	16 (8)	
Transmitted Information	.98 (na)	.96 (.71)	.52 (.26)	.09 (.03)	
Percent of Information	49	49 (46)	27 (13)	04 (02)	
Chi Square	1034 (na)	1006 (746)	531 (253)	79 (29)	
Off Diagonal Chi Square	50 (na)	57 (24)	49 (41)	30 (32)	

* Not Applicable

accuracy, and the lowest variation in accuracy among symbols. For the three manufacturers that had signs with flashing modes, accuracy was considerably reduced when the symbols were presented in the flashing mode.

The next two lines below the average accuracy data in Table 3 show the results of calculations of the information transmitted by each sign. The transmitted information values were calculated from standard bivariate Information Theory formulae (Kantowitz and Sorkin, 1983). The transmitted information values in the table give the amount of information, in bits, received by observers, relative to the total amount of information that the observers could receive under perfect viewing conditions. The amount of information transmitted by each sign shows that the Dambach and Sylvia signs transmitted more information than the Futurit and Yvan Paque signs.

Since the amount of information transmitted is independent of accuracy, perfect information transmission could be obtained even if observers consistently misidentified one set of symbols for another. However, since the transmitted information results for the symbols were consistent with the accuracy results, we can conclude that there weren't any outstanding confusions among the symbols, or outstanding misinterpretations between symbols.

The percent of information transmitted is also presented in the table. This statistic is often called the uncertainty coefficient. It is a measure of how much uncertainty is reduced given that we know what stimulus is being shown to the observer. For the Dambach and Sylvia signs, this is about 49%. In other words, given that we know that the Dambach sign is being viewed, and we know that the left arrow is being shown, our uncertainty about the observer's response is reduced by about half. However, for the Yvan Paque sign, uncertainly is reduced by only about 4%. Clearly the Yvan Paque symbols are not very useful for conveying information.

Below the transmitted information results are chi square statistics that, like the information statistics, were calculated from the confusion matrices in Table 2. The chi square tells, in this case, the degree of relationship between which symbols were presented and which symbols were identified by the observers. The larger the value of the chi square statistic, the stronger is

the relationship between the stimulus events and the response events. A statistically significant chi square means that a real relationship exists. The sizes of the chi squares are consistent with the accuracy results; highest for the Dambach and Sylvia signs, and lowest for the Futurit and Yvan Paque signs. All of the chi square values in Table 3 are significant at the 0.01 probability level; showing that even for the sign on which accuracy was lowest, there was still a statistically significant relationship between the observers' responses and the symbols that were presented. In other words, the relationships between the symbols presented and observers' responses were not random.

The last row in Table 3 shows chi squares calculated for the values in the off-diagonal cells of the Table 2 confusion matrices. These off-diagonal cells contain the error data representing how often observers identified any particular symbol when another particular symbol was actually presented. The chi squares for the error data were calculated to determine whether or not there were systematic error confusions among different symbols. All of the chi square values were significant at at least the .05 level for the off-diagonal chi squares. Although the information analysis indicated that there were no outstanding confusions between or among the symbols, the significant off-diagonal chi squares show that there were statistically significant systematic confusions.

Systematic errors in identifying visual stimuli, such as the lane control symbols, can result from either of two causes. One cause is visual confusions among the symbols. For example, if two of the lane control symbols look alike, there will be a stronger tendency to mistake one of these symbols for the other than if the symbols were not visually confused. The other cause of systematic confusion errors is called response bias. In the case of the lane control symbols, response bias showed up in the aggregated confusion matrix in Table 2 as a strong tendency for observers to report seeing the left and right arrows more often than the down arrow and the X, even though all four symbols were presented equally often.

To separate the effects of visual confusion and response bias in the symbol legibility data, we computed confusion matrices corrected for response bias. The computation was based on using the row and column sums in each

matrix to compute the confusion matrices that would have been produced if observers had reported each symbol equally often. Table 4 shows the results of these computations. The totals in the computed tables do not always match those in the original tables because the computation formulae rounded the results to the nearest whole numbers. The left hand set of sub-tables in Table 4 show the original confusion matrices for the steady state signs. The sub-table at the bottom left shows an aggregate matrix made up of the sum of the four individual matrices.

The set of sub-tables down the centre of the table show the original confusion matrices after being corrected for response bias. The bottom centre sub-table is the corrected aggregate matrix. Notice that the row sums in the corrected matrices are all equal to one another. The sub-tables on the right are simply the corrected matrices collapsed across the diagonals. These diagonal matrices show the total number of times each pair of symbols was confused, one with the other, irrespective of which one was the presented symbol and which one was the response.

The overall effects of correcting for response bias can be seen in the aggregate matrices. The diagonal cells of the original matrix show substantially higher accuracy for the left and right arrows than for the down arrow and the X sign. Although these accuracy differences still appear in the corrected matrix, they are considerably attenuated. In other words, the accuracy differences that appear in the original matrix are due more to response bias effects than they are due to differences in the visibilities of the symbols themselves.

Considering the matrices for the individual signs, all of them except the Sylvia sign show the same response bias trends, and the same symbol accuracy trends. In the case of the Sylvia sign, there did not appear to be a bias against the X as there was in the other signs. In the Sylvia sign, the X was the most accurately identified symbol. During the trials, observers commented that the Sylvia X was considerably brighter than the other symbols. Thus the higher accuracy for this symbol was probably due to at least its higher apparent brightness, if not its higher luminance. Since we have no luminance specifications for the signs, we do not know if the Sylvia X actually did have higher luminance than any of the other signs.

Table 4: Confusion Matrices for the Symbols: Corrected for Response Bias

Sylvia

Original Matrix — Symbol Displayed

Response	Left	Rt	Dwn	X	Sum
Left	119	13	20	11	163
Right	28	130	21	4	183
Down	6	8	112	3	129
X	7	9	7	142	165
Sum	160	160	160	160	640

Corrected for Response Bias — Symbol Displayed

	Left	Rt	Dwn	X	Sum
Left	118	12	19	10	160
Right	22	124	15	-2	160
Down	14	16	120	11	160
X	6	8	6	141	160
Sum	160	160	160	160	640

Collapsed Across Diagonal

	Left	Rt	Dwn	X
Left	118	35	33	16
Right		124	31	6
Down			120	17
X				141

Futurit

Original Matrix — Symbol Displayed

Response	Left	Rt	Dwn	X	Sum
Left	124	17	21	38	200
Right	18	118	20	28	184
Down	4	14	103	31	152
X	14	11	16	63	104
Sum	160	160	160	160	640

Corrected for Response Bias — Symbol Displayed

	Left	Rt	Dwn	X	Sum
Left	114	7	11	28	160
Right	12	112	14	22	160
Down	6	16	105	33	160
X	28	25	30	77	160
Sum	160	160	160	160	640

Collapsed Across Diagonal

	Left	Rt	Dwn	X
Left	114	19	17	56
Right		112	30	47
Down			105	63
X				77

Yvan Paque

Original Matrix — Symbol Displayed

Response	Left	Rt	Dwn	X	Sum
Left	77	39	31	34	181
Right	27	56	30	50	163
Down	23	39	75	45	182
X	29	22	20	27	98
Sum	156	156	156	156	624

Corrected for Response Bias — Symbol Displayed

	Left	Rt	Dwn	X	Sum
Left	71	33	25	28	156
Right	25	54	28	48	156
Down	17	33	69	39	156
X	44	37	35	42	156
Sum	156	156	156	156	624

Collapsed Across Diagonal

	Left	Rt	Dwn	X
Left	71	58	41	71
Right		54	61	85
Down			69	73
X				42

Dambach

Original Matrix — Symbol Displayed

Response	Left	Rt	Dwn	X	Sum
Left	138	15	19	25	197
Right	15	136	14	14	179
Down	3	2	119	6	130
X	4	7	8	115	134
Sum	160	160	160	160	640

Corrected for Response Bias — Symbol Displayed

	Left	Rt	Dwn	X	Sum
Left	129	6	10	16	160
Right	10	131	9	9	160
Down	11	10	127	14	160
X	11	14	15	122	160
Sum	160	160	160	160	640

Collapsed Across Diagonal

	Left	Rt	Dwn	X
Left	129	16	20	26
Right		131	19	23
Down			127	28
X				122

Aggregate

Original Matrix — Symbol Displayed

Response	Left	Rt	Dwn	X	Sum
Left	458	84	91	108	741
Right	88	440	85	96	709
Down	36	63	409	85	593
X	54	49	51	347	501
Sum	636	636	636	636	2544

Corrected for Response Bias — Symbol Displayed

	Left	Rt	Dwn	X	Sum
Left	432	58	65	82	636
Right	70	422	67	78	636
Down	47	74	420	96	636
X	88	83	85	381	636
Sum	636	636	636	636	2544

Collapsed Across Diagonal

	Left	Rt	Dwn	X
Left	432	128	112	170
Right		422	141	161
Down			420	181
X				381

From the steady mode accuracy data given in Table 3, accuracy was highest for the Sylvia and Dambach signs. Both signs produced the same average accuracies. However, the two signs produced rather different patterns of accuracies for individual symbols. And when the accuracy results were corrected for response bias, the two signs produced yet different patterns of accuracies for individual symbols. These different patterns are illustrated in Figure 7. Figure 7 shows a bar graph representing the identification accuracies for the symbols in each of the two signs, for the original uncorrected accuracies and for the accuracies after correcting for response bias. It can be seen that when the Dambach accuracies are corrected for response bias, the four Dambach symbols all produce about the same accuracy. In the case of the Sylvia symbols, the correction for response bias yields about equal accuracies for the left, right, and down arrows, but still leaves the X with substantially higher accuracy than any of the other three symbols.

From an Information Theory perspective, and other things being equal, it is best to have a set of symbols in which the symbols are equally legible, and equally distinguishable from one another. This suggests that, for the steady state mode, the Dambach sign gave the best performance. The confusion matrices on the right of Table 4 also support this conclusion. For maximum recognition of a set of symbols, all of the symbols should be equally distinguishable from one another. For the Dambach sign, the confusions among the different pairs of symbols are all about the same, ranging from 16 confusions between the left and right arrows to 28 confusions between the down arrow and the X. In the Sylvia sign, the confusions are more variable, ranging from six confusions between the right arrow and the X, to 35 confusions between the left and right arrows.

Analysis of the results for the symbol signs in flashing mode showed that the flashing mode accuracy produced the same pattern of results as the steady mode except for a reduction in accuracy. A regression analysis in which the steady mode percent accuracies were used to predict the flashing mode percent accuracies showed that there was a correlation of 0.92 between steady mode accuracies and flashing mode accuracies. Since the Dambach sign had no flashing mode, the regression analysis included only data from the Sylvia, Futurit, and Yvan Paque signs. The regression equation showed that flashing

mode accuracy could be predicted by the equation;

$$F = -3.9 + 0.84S$$

where F is flashing mode accuracy, and S is steady mode accuracy. The equation shows that flashing mode reduces accuracy by about 16 percent, plus a further constant reduction of about four percent.

Because of the high correlation between flashing mode accuracy and steady mode accuracy, the detailed comparisons of signs shown in Table 4 for the steady mode would show essentially the same results for the flashing mode. However, since the Dambach sign had no flashing mode, the Sylvia sign would emerge as the sign that generated the highest accuracy results for the flashing mode. Given the high correlation between steady mode accuracy and flashing mode accuracy, it is likely that the Dambach sign would have produced flashing mode results comparable with its steady mode results. However, this remains a matter of speculation.

5.1.2 Numbers Legibility

Table 5 shows the confusion matrices for the numeric characters. Table 5 is equivalent to Table 2 which showed the confusion matrices for the symbols. The bottom right matrix is an aggregate matrix located where a flashing mode matrix would have appeared for the Dambach sign, if the Dambach sign had a flashing mode. Notice also that, unlike the other signs, the Dambach sign did not have any 60 or 70 presentations, although observers did give 60 or 70 as responses.

The data in Table 5 show characteristics similar to those already seen in Table 2. The diagonal cells show accuracies generally in the region of 50 percent. The aggregate matrix again shows a response bias effect; in this case, a bias favoring the 30, 50 and 60 displays.

Table 6 shows a summary of the accuracy analysis for the numbers, similar to the summary shown in Table 3 for the symbols. In the steady mode the Dambach, Sylvia, and Yvan Paque signs all gave about the same average accuracy; just under 50 percent. Accuracy on the Futurit sign was only just above chance. Since the accuracy calculations were corrected for guessing,

Table 5: Confusion Matrices for the Numbers

STEADY MODE
Number Displayed

Responses

SYLVIA

	30	50	60	70	80	90	100	SUM
30	103	15	8	13	17	25	3	184
50	20	107	14	10	25	32	4	212
60	7	16	124	7	29	13	3	199
70	17	3	3	114	11	11	3	162
80	5	10	5	2	59	17	6	104
90	6	9	3	12	18	60	2	110
100	2	0	3	2	1	2	139	149
SUM	160	160	160	160	160	160	160	1120

FUTURIT

	30	50	60	70	80	90	100	SUM
30	30	32	27	26	20	14	19	168
50	32	33	28	30	31	30	22	206
60	23	30	43	27	41	26	25	215
70	23	19	13	29	9	10	14	117
80	16	15	22	13	22	18	19	125
90	18	18	16	15	16	52	9	144
100	18	13	11	20	21	10	52	145
SUM	160	160	160	160	160	160	160	1120

YVAN PAQUE

	30	50	60	70	80	90	100	SUM
30	108	6	14	16	4	9	19	176
50	13	86	23	8	10	19	15	174
60	6	31	79	5	31	9	13	174
70	9	6	10	98	3	9	5	140
80	11	13	16	4	104	7	6	161
90	11	16	10	22	6	103	4	172
100	2	2	8	7	2	4	98	123
SUM	160	160	160	160	160	160	160	1120

DAMBACH

	30	50	60	70	80	90	100	SUM
30	113	10	0	0	3	5	18	149
50	12	128	0	0	10	6	10	166
60	4	9	0	0	27	1	14	55
70	10	1	0	0	4	1	2	18
80	11	5	0	0	98	13	3	130
90	6	5	0	0	10	134	5	160
100	4	2	0	0	8	0	107	121
SUM	160	160	0	0	160	160	159	799

FLASHING MODE
Number Displayed

SYLVIA

30	50	60	70	80	90	100	SUM
75	16	7	18	13	17	2	148
27	95	21	16	25	37	9	230
15	17	96	8	34	23	4	197
15	8	4	99	12	8	5	151
13	9	19	6	47	31	6	131
12	13	9	9	25	40	1	109
3	2	4	4	4	4	133	154
160	160	160	160	160	160	160	1120

FUTURIT

30	50	60	70	80	90	100	SUM
25	27	25	27	20	17	20	161
23	29	29	33	29	28	15	186
28	41	31	35	29	32	31	227
28	18	22	22	13	17	19	139
12	21	13	14	20	19	21	120
22	8	17	10	21	29	19	126
22	16	23	19	28	18	35	161
160	160	160	160	160	160	160	1120

YVAN PAQUE

30	50	60	70	80	90	100	SUM
93	7	18	25	8	12	24	187
15	65	21	16	10	11	12	150
9	27	61	13	33	13	9	165
18	12	14	67	4	13	12	140
14	21	13	10	80	14	7	159
10	19	9	15	15	94	8	170
1	9	24	14	10	3	88	149
160	160	160	160	160	160	160	1120

SUMS ACROSS ALL SIGNS

30	50	60	70	80	90	100	SUM
547	113	99	125	85	99	105	1173
142	543	136	113	140	163	87	1324
92	171	434	95	224	117	99	1232
120	67	66	429	56	69	60	867
82	94	88	49	430	119	68	930
85	88	64	83	111	512	48	991
52	44	73	66	74	41	652	1002
1120	1120	960	960	1120	1120	1119	7519

Table 6: Comparisons of Manufacturer's Signs: Accuracy, Information and Chi Squares for the Numbers
(Numbers in Brackets are for Flashing Characters)

	MANUFACTURER				
Numbers	Dambach	Sylvia	Futurit	Yvan Paque	Average
30	51 (na)	50 (33)	5 (2)	54 (44)	40 (26)
50	60 (na)	53 (45)	7 (4)	40 (27)	40 (25)
60	na (na)	64 (46)	13 (5)	35 (24)	37 (25)
70	na (na)	57 (48)	4 (0)	47 (28)	36 (25)
80	41 (na)	23 (15)	0 (-1)	51 (36)	29 (28)
90	41 (na)	24 (11)	19 (4)	50 (45)	34 (20)
100	47 (na)	73 (69)	19 (8)	47 (41)	47 (39)
Average Accuracy	48 (na)	49 (38)	10 (3)	46 (35)	38 (25)
Standard Deviation	7 (na)	18 (19)	7 (3)	6 (8)	
Transmitted Information	1.20	1.07(.75)	.09 (.04)	.92 (.58)	
Percent of Information	46	39 (27)	3 (1)	33 (21)	
Chi Square	1702 (na)	2409 (1654)	168 (56)	2101 (1207)	
Off Diagonal Chi Square	275 (na)	273 (268)	112 (94)	195 (129)	

the accuracy values take account of the fact that observers had only a one in seven chance of correctly guessing Sylvia, Futurit, or Yvan Paque displays, compared with a one in five chance of correctly guessing the Dambach display. As the standard deviations show, the Yvan Paque sign showed the least variability among accuracies for individual numbers. The high variability shown for the Sylvia sign was almost entirely due to the low accuracy on 80 and 90 compared with quite high accuracy on the other numbers. The Dambach sign gave fairly consistent accuracies across the five numbers that it could display.

Accuracies in the flashing mode were, like the flashing symbol accuracies, about 20 percent lower than accuracies in the steady mode. Similarly, accuracy differences among numbers in the flashing mode were consistent with the differences among numbers in the steady mode.

As in the case of the symbol signs, the Information Theory measures show results consistent with the accuracy results. Thus the number signs did not produce any outstanding confusions or display misinterpretations that would have caused the information results to be inconsistent with the accuracy results.

The chi square results show that there were significant relationships between which numbers were displayed and which numbers observers identified; even for the Futurit sign on which accuracy was quite low. The off-diagonal chi squares also show that there were statistically significant confusions among particular number displays.

The confusion matrices, corrected for response bias are presented in Table 7. A corrected table for the Dambach sign that would be comparable with the tables for the other three signs could not be produced because of the lack of the 60 and 70 displays. Like the analysis for the symbols, correcting for response bias did not significantly alter the pattern of results for the uncorrected data, as shown in Table 5. In particular, correcting for response bias did not change the low accuracy found for the 80 and 90 displays in the Sylvia sign. For the purposes of choosing the sign that produced the best overall performance, the Yvan Paque sign appeared to be the best choice.

Corrected for Response Bias

Sylvia

Response	Symbol displayed 30	50	60	70	80	90	100	Sum
30	100	12	5	10	14	22	0	160
50	13	100	7	3	18	25	-3	160
60	1	10	118	1	23	7	-3	160
70	17	3	3	114	11	11	3	160
80	13	18	13	10	67	25	14	160
90	13	16	10	19	25	67	9	160
100	4	2	5	4	3	4	141	160
Sum	160	160	160	160	160	160	160	1120

Collapsed Across Diagonal

	Symbol displayed 30	50	60	70	80	90	100
30	100	24	6	26	27	35	3
50		100	17	5	36	41	-2
60			118	4	36	18	2
70				114	21	30	6
80					67	50	17
90						67	13
100							141

Futurit

Response	Symbol displayed 30	50	60	70	80	90	100	Sum
30	29	31	26	25	19	13	18	160
50	25	26	21	23	24	23	15	160
60	15	22	35	19	33	18	17	160
70	29	25	19	35	15	16	20	160
80	21	20	27	18	27	23	24	160
90	20	20	18	17	18	54	11	160
100	20	15	13	22	23	12	54	160
Sum	160	160	160	160	160	160	160	1120

	Symbol displayed 30	50	60	70	80	90	100
30	29	56	41	54	40	33	38
50		26	44	49	44	44	31
60			35	38	60	36	30
70				35	33	33	42
80					27	41	47
90						54	23
100							54

Yvan Paque

Response	Symbol displayed 30	50	60	70	80	90	100	Sum
30	106	4	12	14	2	7	17	160
50	11	84	21	6	8	17	13	160
60	4	29	77	3	29	7	11	160
70	12	9	13	101	6	12	8	160
80	11	13	16	4	104	7	6	160
90	9	14	8	20	4	101	2	160
100	7	7	13	12	7	9	103	160
Sum	160	160	160	160	160	160	160	1120

	Symbol displayed 30	50	60	70	80	90	100
30	106	15	16	26	13	16	24
50		84	50	15	21	31	20
60			77	16	45	15	24
70				101	10	32	20
80					104	11	13
90						101	12
100							103

Table 7: Confusion Matrices for the Numbers Corrected for Response Bias

Although the Yvan Paque sign gave average accuracy two or three percentage points lower than the average accuracy on either the Sylvia or Dambach signs, consideration of other factors more than compensate for the slightly lower average accuracy of the Yvan Paque. First, although the Dambach sign had the highest average accuracy, it did not display the 50 and 60 numbers. As the data for both symbols and numbers show, high average accuracy for a sign's displays is no guarantee of high accuracy on each and every one of the sign's characters. Thus the high accuracy on the numbers that the Dambach could display, provides no basis for expecting that there would also have been high accuracy for the 50 and 60 displays had the sign been capable of displaying them.

In the case of the Sylvia sign, five of the numeric displays produced quite high accuracy while the two remaining numeric displays produced quite low accuracy. In contrast, the Yvan Paque sign produced nearly the same accuracy on every display. Figure 8 shows a bar graph comparing display accuracies for individual number displays, across the Sylvia, Yvan Paque, and Dambach signs. Futurit accuracies were less than 20% and are not illustrated. The graph shows that accuracy was relatively constant across number displays for the Yvan Paque sign, and rather variable across displays on the Sylvia sign. The accuracies were constant across the displays that the Dambach sign supported. However, there is no way of knowing what the accuracies might have been for the 60 and 70 displays that Dambach did not support.

A review of the confusion matrices to see if there were any pairs of number displays that were outstandingly more often confused than other pairs showed that confusions were fairly evenly distributed among possible pairs for all three signs. For the Sylvia sign, there were relatively high numbers of confusions associated with the 80 and 90 displays. However, these confusions were due more to the low legibilities of the 80 and 90 displays than they were due to 80 or 90 displays being particularly confused with other specific displays. When 80 or 90 was presented on the Sylvia sign, one number was as likely to be given as the error response as any other number.

Analysis of the results for the number signs in flashing mode showed that the flashing mode again produced the same pattern of results as the steady mode except for a reduction in accuracy. A regression analysis in which the

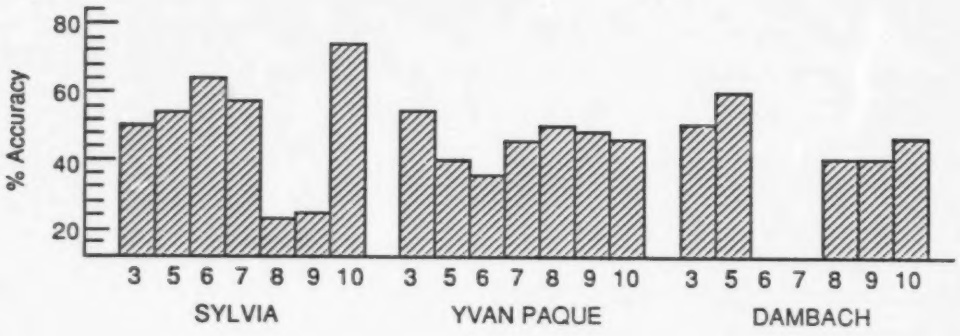

Figure 8/ Accuracies for Sylvia, Yvan Paque and Dambach Numbers (Table 6)

steady mode percent accuracies were used to predict the flashing mode percent accuracies showed a correlation of 0.97 between steady mode accuracies and flashing mode accuracies. Since the Dambach sign had no flashing mode, the regression analysis included only data from the Sylvia, Futurit, and Yvan Paque signs. The regression equation showed that flashing mode accuracy could be predicted by the equation;

$$F = -6.0 + 0.89S$$

where F is flashing mode accuracy, and S is steady mode accuracy. The equation for numbers has slightly different coefficients than the one for symbols; but the net effect is still a 15 to 20 percent reduction in accuracy for the flashing mode.

As for the symbols, because of the high correlation between flashing mode accuracy and steady mode accuracy, the detailed comparisons of signs shown in Table 5 and Figure 8 for the steady mode would show essentially the same results for the flashing mode. Consequently, the conclusion that the Yvan Paque sign produced the best overall performance would hold for the flashing mode as well as for the steady mode. Again, because of the high correlation between steady mode accuracy and flashing mode accuracy, it is likely that the Dambach sign would have produced flashing mode results comparable with its steady mode results. However, since the Dambach sign had no flashing mode, this again remains a matter of speculation.

5.2 Comprehension

This section describes the results of asking observers to describe what they thought the different sign displays meant. The first signs that the observers were asked to identify were the 30 and 100 speed control signs. Of the 40 observers, 38 said that the 30 represented a speed limit. One person thought that it was too low a number to be a speed limit, and another thought that it meant that something was happening, or would happen, 30 metres ahead.

The colour of the 30 sign varied from sign to sign. The colour of the 30 was white on the Futurit sign, yellow/amber on the Sylvia and Yvan Paque signs, and almost orange on the Dambach sign. Twenty people mentioned the

differences in colour among the signs, noting that the colours added nothing to the meaning of the sign. Two people mentioned that the orange and yellow signs 'stood out more'.

All but one observer thought that the 100 represented a speed limit sign. The same person who thought that the 30 was a distance measure, also thought that the 100 meant that something would happen 100 feet or metres ahead on the highway.

All of the 100 signs were green. The Sylvia 100 was the brightest. Only nine observers noted the colour; but they also said that the colour did not add anything extra to the meaning of the sign.

After describing what the number signs meant, observers were then asked for the meanings of the symbol signs. Although the symbols were in colour, no one noted or commented on the colours. Eighty five percent of the observers correctly identified the left, right and down arrows as merge left, right, and lane open, respectively. The X, meaning lane closed, was interpreted correctly by only 68% of the observers.

The majority of the observers thought that the down arrow indicated that the driver should stay in the lane, or that the lane was open or clear ahead. Other opinions were;

- danger, proceed with caution because there is something ahead in this lane (2 people);
- a valley ahead (one person);
- use this lane to exit to a street ahead (one person);
- there is another sign ahead that will give further directions (one person); and
- turn here (one person).

Thirty four observers correctly identified the right facing arrow as an indication to merge right, four of whom thought it indicated a detour to the right because of an accident ahead. Other opinions were;

- an off-ramp to the right (three people);
- the lane turns to the right (two people); and
- the right lane must exit (one person).

The left facing arrow elicited similar responses, but for the left direction.

Of the signs used in the legibility study, the X elicited the fewest number of correct guesses. Only 27 people thought that it indicated that the driver was not to use the lane, or that the lane was closed. Other responses were;

- stop (five people);
- a crosswalk or crossing of some kind (four people);
- slow down, proceed with caution (three people); and
- railway tracks (one person).

The symbol for road construction was shown to 30 observers. Dambach was the only manufacturer who attempted to produce this symbol. Only two people guessed correctly, that it meant 'men at work'. Eleven people said they had no idea what it meant. The other guesses were;

- stop, the lane is closed, do not enter (11 people);
- caution, slow down or yield (four people);
- gravel road ahead (one person); and
- danger ahead (one person).

5.3 Preference

Figure 9 shows which numbers and symbols were preferred by the observers as the best way to convey the intended messages. As described earlier, observers ranked the sign displays according to preference. This was done by showing, for example, the four 30 signs and asking the observer to rank the four 30 signs from most preferred to least preferred. The overall preference ratings were determined by calculating the observers' median ranking for each display.

Observers generally preferred the number representations on the Yvan Paque and Dambach signs. These numbers are composed of light sources that give characters the appearance of being made up of continuous line strokes; while the less preferred Sylvia and Futurit numbers are clearly in a dot

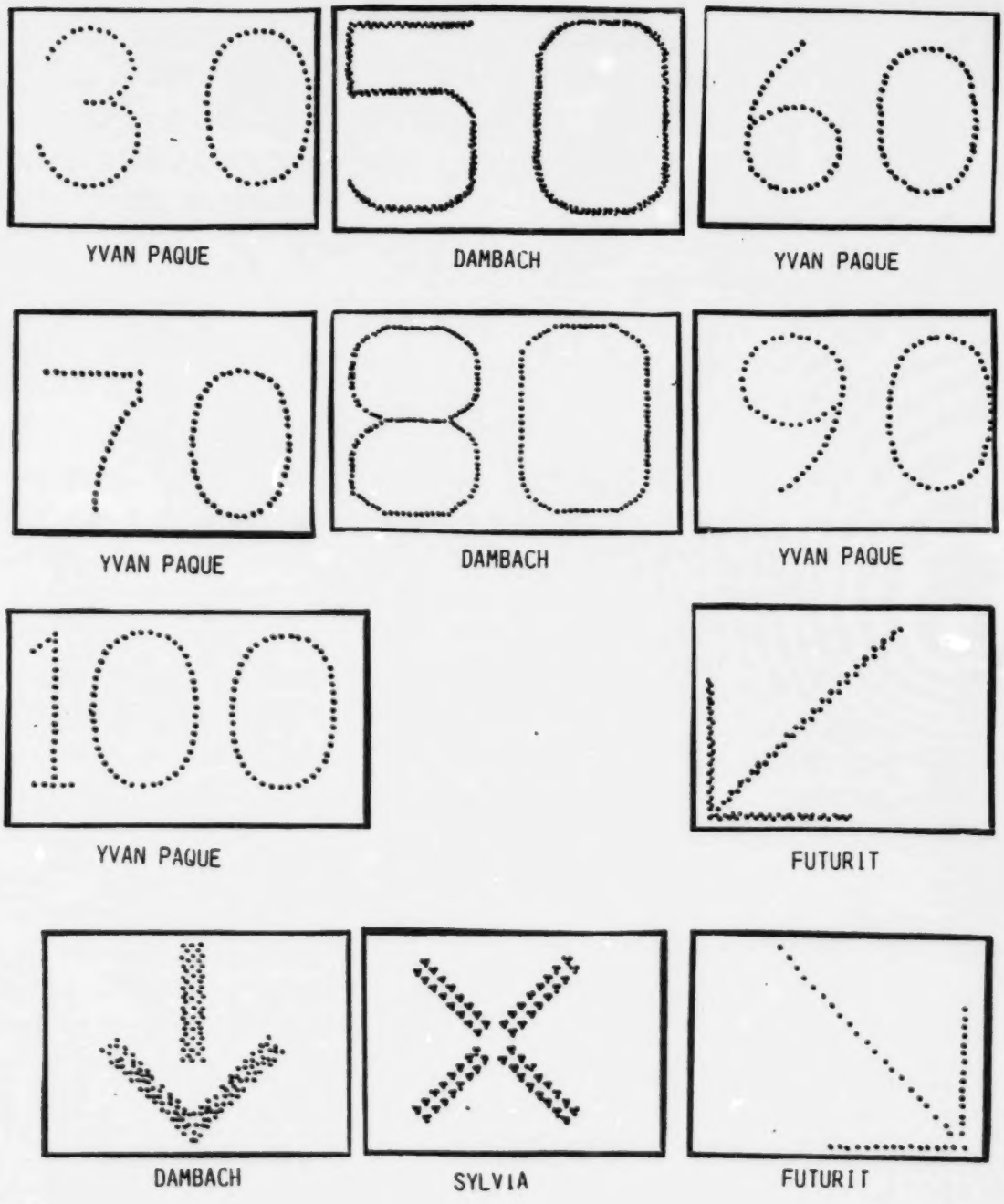

Figure 9/ Number and Symbol Preferences

Table 8: Sign Displays Ordered According to Accuracy

	Display			% Accuracy
Symbol Signs	Sylvia	Red X		64
	Dambach	Left Arrow		61
	Dambach	Right Arrow		60
	Sylvia	Right Arrow		56
	Futurit	Left Arrow		53
	Dambach	Down Arrow		49
	Sylvia	Left Arrow		49
	Futurit	Right Arrow		49
	Dambach	Red X		47
	Sylvia	Down Arrow		45
	Futurit	Down Arrow		39
	Yvan Paque	Down Arrow		34
	Yvan Paque	Left Arrow		24
	Futurit	Red X		14
	Yvan Paque	Right Arrow		11
	Yvan Paque	Red X		-8 *
Number Signs	Sylvia		100	73
	Sylvia		60	64
	Dambach		50	60
	Sylvia		70	57
	Yvan Paque		30	54
	Sylvia		50	53
	Dambach		30	51
	Yvan Paque		80	51
	Sylvia		30	50
	Yvan Paque		90	50
	Yvan Paque		100	47
	Yvan Paque		70	47
	Dambach		100	47
	Dambach		80	41
	Dambach		90	41
	Yvan Paque		50	40
	Yvan Paque		60	35
	Sylvia		90	24
	Sylvia		80	23
	Futurit		100	19
	Futurit		90	19
	Futurit		60	13
	Futurit		50	7
	Futurit		30	5
	Futurit		70	4
	Dambach		60	na
	Futurit		80	0
	Dambach		70	na

* Negative value means accuracy less than chance

matrix style. The preference for a continuous line stroke style carries over to the left and right arrows, but not to the down arrow or the X; where the preference was for bold, or thick stroke, representations. Thus, Futurit had the most preferred left and right arrows; Dambach the most preferred down arrow, and Sylvia the most preferred X.

A comparison of the preference ratings with the accuracy data showed a significant correlation of 0.74 between accuracy and preference. It would appear that the legibility of a sign is a substantial factor in determining how well the sign is perceived to convey its intended message. Since observers were asked to give their preference ratings before the legibility trials were run, their preferences could not have been influenced by their knowledge of their performance on the legibility trials.

5.4 Characteristics of Legible Signs

This section describes analyses done to see if there were particular design features that tended to make some displays more legible than others. Table 8 shows a listing of the individual sign displays ordered according to legibility. The top part of the table shows the symbol displays and the bottom part of the table shows the number displays. The percent accuracy values used to order the displays are the accuracy values for the steady mode given in Tables 3 and 6.

Table 8 shows that no one manufacturer had a monopoly on producing consistently legible signs. For the symbol signs, the most legible displays were made of solid, thick stroke font styles. Thus the Dambach symbols, and the Sylvia X appear at the top of the symbols list. A set of number displays made up from the seven most legible number displays would consist of the Yvan Paque 30, the Dambach 50, the Sylvia 60, the Sylvia 70, the Yvan Paque 80, the Yvan Paque 90, and the Sylvia 100. Inspection of these displays will show that they would produce a rather odd assortment of number displays to put on road signs.

The sign displays varied from one to another on several dimensions of specific design features such as size of character, colour, and font style.

These characteristics did not vary systematically within the set of symbols and numbers that were presented to observers during the trial. However, there were enough coincidental variations to make it worth trying to see if there were certain common features in the designs of the displays that produced higher legibility.

For the purposes of this analysis, we tabulated for each display the following characteristics and quantities;

- the height of the display characters;
- the width of the display characters;
- the colour of the display (white, green, red, or amber); and
- font style (line stroke versus dot matrix).

We did not include steady versus flashing mode as a variable in this analysis since we already knew that the flashing mode reduces accuracy by 15 to 20 percent.

We then computed regressions using these variables to predict accuracy. There was a small positive relationship between accuracy and display height, display width, and colour. There was no relationship between accuracy and font style, despite the previous finding that preference seemed to be related to both font style and accuracy. Although there was a small relationship between accuracy and height and width, the relationship was not statistically significant. Although there were variations in the sizes of the display characters, the variations were not large enough to significantly affect the accuracy results.

There was a statistically significant relationship between accuracy and display colour. This relationship turned out to be due mainly to whether or not the display colour was amber. This is consistent with the fact that a number of observers mentioned that the amber displays seemed to be brighter than the other displays. In fact, although we had no luminance data for the signs, the colour affect found here should probably be treated as a luminance effect rather than a colour effect.

The same regression was performed to include only data from the number displays. The equation yielded different results from those produced when

both the symbol and number displays were in the analysis. In this case, the regression showed significant effects due to height, width, and font style. The regression equation produced from this analysis is given by

$$L = 0.61H - 0.59W + 0.56F$$

where L is legibility, H is height, W is width, and F is font style. Font style was coded as 1 for line stroke font, and 0 for dot matrix font. The coefficients in the equation are the standardized coefficients to show the relative weights of the variables independent of the units in which the were measured. All three regression coefficients were statistically significant. The regression coefficients show that height, width, and font style each contributed about equally to variation in legibility. The multiple correlation between legibility and the three independent variables was 0.78. The square of the multiple correlation shows that altogether, character height, width, and font style produced just over 60 percent of the variation in legibiiity.

The positive relationship between height and legibility shows that, for the number displays, even with the small differences among the heights of the characters in the different displays, displays having taller characters yielded higher legibility. The negative relationship between legibility and width probably reflects the fact that the more legible displays contained characters that had optimum height to width ratios, while the less legible displays had character widths that were too wide relative to their heights. Character legibility is usually a function of character height, and the degree to which the average character's height to width ratio is optimum. An average height to width ratio of about 5:3 is close to optimum for typical fonts. Making a character too wide or too narrow, relative to its height, decreases legibility.

The effect of font style in the regression equation showed that the line stroke font style produced higher legibility than the dot matrix style characters. The effect of font style on legibility is consistent with the preference data and the raw legibility data. Observers preferred the line stroke fonts, and these were also the displays that were recognized most accurately. The effect of font style was reversed for the symbol displays, where symbols made in a line stroke font yielded the lowest accuracy. The most accurately recognized symbols were produced in a thick solid stroke style

(see the Dambach symbols, and the Sylvia X).

Taken together, the analysis of display features suggest that the following general design features contribute to the legibility of number displays;

- maximizing character height;
- using proper character height to width ratios; and
- using a line stroke font style.

For the number signs a line stroke style amounted to making characters out of single lines of light sources, with the light sources being spaced closely enough together to give the appearance of a single continuous line.

The features that produced legible symbols were not so obvious as for the number displays. From the analysis, it appeared that producing the symbols to appear as solid geometric forms yielded the best results.

The results of the regression analyses should not be treated as a definitive analysis of the design features that produced the highest legibility. The features that were analyzed here did not vary systematically among the signs. Thus the correlations among the features represented more or less random occurrences of whatever combinations of features happened to occur in the same displays. For example, the 100 sign was always green; while green was never used for any of the other number signs. In a proper analysis of colour effects, each display would have to be tested with each of the colours. Thus the analysis presented here represents a precaution against overlooking any design features that might have had an unexpected pronounced effect on the legibility results.

6.0 DISCUSSION

In this section, we will first discuss conclusions about which manufacturers signs provided the highest overall legibility. Then we will discuss conclusions about display design features that influence legibility. At the end of the discussion, we will discuss the results of the field test by reference to past work on legibility.

For the symbol displays, the Dambach symbols provided the best compromise among factors that influenced legibility. The Dambach symbols were recognized with relatively high accuracy, and they yielded consistent accuracy across displays. While the Sylvia displays produced relatively high average accuracy, accuracy was inconsistent from symbol to symbol. From the perspective of information transmitted, it is important that displays have not only high average legibility, but also consistent legibility across symbols. For this reason, the Dambach signs would be the best choice for a set of symbol signs.

In the case of the number displays, Dambach, Yvan Paque, and Sylvia all produced relatively high average legibility. However, only the Yvan Paque displays produced consistent legibility across displays. In the Sylvia displays, the 80 and 90 displays had very low legibilities, which had the net effect of causing observers to randomly guess from among 30, 50, 60, 70, and 80 or 90 whenever one of 80 or 90 displays was shown. In the Dambach displays, the displays that were implemented had consistently high legibility, but the lack of 60 and 70 displays makes it impossible to know if this consistency would extend to the 60 and 70 displays. As seen in the Sylvia number displays, the high legibility of five characters was no guarantee of high legibility on two other characters. Taking into account both average legibility and consistency of legibility, the Yvan Paque number displays appear to be the best choice for the number displays.

Since the best choice of manufacturer for the symbol signs is different from the best choice of manufacturer for the number displays, the next question is which of Dambach or Yvan Paque provides the best compromise in a choice of one manufacturer for both the symbol and number displays. The only way in which Yvan Paque could serve as a suitable choice would be if Yvan Paque could implement the symbol displays as solid symbol displays similar to the Dambach symbols, and the Sylvia X. The line stroke font style, like that used for the Yvan Paque symbols, yielded low legibility on all displays where it was used. To provide symbol legibility consistent with the Dambach signs, the Yvan Paque would have to display solid figure symbols similar to the Dambach symbols and the Sylvia X.

Turning to Dambach, both the Dambach symbol displays and number displays yielded consistently high legibility performance on all the displays that were implemented by Dambach. However, the lack of 60 and 70 displays makes it impossible to be certain if the legibility performance on 60 and 70 displays, if they were created, would be the same as the legibility performance on the other Dambach displays. On the other hand, the fact that the line stroke font styles reliably produced the highest legibility for numbers suggests that if Dambach implemented 60 and 70 displays with this font style, the Dambach would be a reasonable single choice for all of the signs.

We note that the Dambach signs had no flashing mode. However, the high correlation between steady mode and flashing mode accuracy indicates that if Dambach signs had had a flashing mode, the Dambach flashing mode legibility results would have been the same as for the steady mode, except for a proportionate decrease in legibility produced by the flashing mode for all signs.

In addition to providing information for comparing manufacturers' signs, the field trial provided some information about design features. The analysis of design features showed that maximizing character height, and having proper character height to width ratios influenced the legibility of the number signs. In addition, the use of a line stroke font style was superior to a dot matrix font style. For symbols, the use of a solid geometric form appeared to be superior to a line stroke style.

The results showed that the flashing sign mode produced a substantial 15 to 20 percent reduction in legibility. This is consistent with the literature on the effects of flashing visual displays. Although flashing displays attract attention, they also decrease recognition accuracy. In general, flashing displays should be restricted for use in attracting attention and should be avoided for use in conveying any other information.

Rules for the use of flashing displays also apply to the use of colour. As a rule, the use of colour in displaying text information decreases legibility rather than increases it. Guidelines for cathode ray tube (CRT) text displays that use colour specify increasing the size or contrast of the text to compensate for the use of colour. The more colours used, the greater

will be the increase in size or contrast needed to compensate for the effect of the colours. As a rule colour should not be used to display text or symbolic information unless there are compelling reasons to do so. If the only purpose of using colour is to increase the redundancy of the information, for example, always displaying the 100 sign in green; the use of colour is likely to be at best self defeating; and at worst, to have an effect opposite to the one intended. Other research suggests that the use of colour is appropriate for warning signals, but not for information displays.

The data from the trial showed that observers' identifications of signs could be influenced by response bias, especially in identifying the symbols. In the case of the symbols, response biases had significant effects on recognition accuracy. In the case of the number signs, although there were response biases evident in the data, the biases did not materially affect recognition accuracy. Normally, response bias effects can be predicted from how familiar an observer is with the display information, and from the observer's expectations. For example, a driver going by a road side sign in poor visibility near Toronto is not likely to interpret the information on a partially seen sign as "10 Km. to Trenton" instead of as "10 Km. to Toronto". For displays of symbols that may be unfamiliar, or be in an unfamiliar context, it is difficult to predict the response biases.

In the case of the symbols used in this trial, there was no basis for predicting the pattern of response bias that was actually found. In addition, the comprehension results showed that observers were not always consistent in their interpretations of what the symbols meant. Experience with designing symbolic signs has shown that it is indeed difficult to design symbols whose meanings are unambiguous. If the information display designer has a choice, alphanumeric text, is better than symbols for conveying information.

In the data analysis, response bias effects were corrected so that the effects of visual confusions could be determined independently of the effects of response bias. It should be kept in mind that this analysis did not correct for response bias in any literal sense. All of the signs, including the most legible ones, will show the same response bias effects in conditions of low visibility. In addition, as visibility gets worse, the greater will be the response bias effects. The only way in which response bias effects can be

avoided is by maximizing the visibility of the information to be displayed. This brings up the question of how the specifications of the message signs compare with specification standards for legible text displays.

At present, the bulk of the literature on visual displays applies to displays seen at relatively short viewing distances, ranging from normal reading distance to distances of several metres. There appears to have been less research done on the legibility of displays viewed at distances similar to the distances at which the message signs were meant to be read.

There are three major variables that determine legibility; character height and height-width ratio, contrast, and font. Of these variables, character height and contrast are the most important. Generally a height-width ratio close to 5:3 will produce adequate legibility. Similarly, a simple font style, such as the one used in the Dambach and Yvan Paque signs will produce good legibility. Although font and proportion are significant factors in legibility, producing characters with poor font or proportion characteristics is more often a matter of lack of knowledge about these matters than it is a matter of technical feasibility.

In the case of height and contrast, there are often economic factors that steer design towards making characters as small as possible, and towards presenting the text with low contrast media. On the other hand, legibility is a directly increasing function of height and contrast. In the case of the message signs, current guidelines for producing minimum legibility suggest that the message signs tested in this field trial are too small. For example, one guideline for printed characters read at short distances (Peters & Adams, 1959) gives a formula for character height as a function of viewing distance as:

$$H = 0.0022D + K1 + K2$$

where H is character height, D is viewing distance, K1 is a correction factor for viewing distance and illumination, and K2 is a correction factor for the importance of the material to be read.

Ignoring the correction factors, the formula is equivalent to saying that the minimum character height for a legible sign corresponds to a visual angle of 7.5 minutes of arc. Studies of legibility of material viewed at different

distances have consistently shown legibility to be a function of visual angle, independent of viewing distance. Thus, for example, if a character is twice as legible as another character at one viewing distance, it would still be twice as legible as the other character at any other viewing distance. Although extrapolations always have to be treated with caution, extrapolations of relative legibility generally produces consistent results. However, extrapolations of absolute legibility results are likely to be less reliable. At long distances, such things as atmospheric interference are bound to have significant effects that would be negligible at short viewing distances. Thus a font having a 7.5 minute of arc character height, and that is just barely legible at 10 metres, will likely have to have a character height of more than 7.5 minutes of arc to be legible at several hundred metres.

Published values of the correction factors are available for only short viewing distances. However, if we ignore the correction factors, the recommended message sign character height would be 0.51m for the 233m viewing distance used in the field trial, or 0.66m at 300m viewing distance which is the maximum distance at which the signs must be legible. The symbols and characters on the signs in the field trial had character heights of about 0.5 metres or less. This means that at the viewing distance used in the field trial, the characters on the signs were only just large enough to be considered adequate for reading print under good viewing conditions. When correction factors are used to account for viewing conditions, the sign characters turn out to be substantially smaller than would be considered adequate for reading printed text. As an example, the following recommended character heights were extrapolated to the viewing distances at which the message signs are to be viewed. The following table gives recommended character heights calculated for a correction factor corresponding to viewing important information.

Viewing Condition	Recommended Character Height in metres at 233m	300m
Favorable (good illumination)	1.04	1.35
Moderate	1.46	1.89
Unfavorable (poor illumination)	1.88	2.43

The recommended character heights are considerably greater than the ones specified for the present signs.

Guidelines for the use of colour lead to similar conclusions (Meister, 1983) about the adequacy of the present sign specifications. This guideline gives character heights for CRT displays as a function of viewing distance and numbers of colours used to display information. The recommended character height as a function of number of colours is

$$H = 15 + (N-1)$$

where H is the character height in minutes of viewing angle, and N is the number of display colours (for N from 1 to 6).

The following table gives, as an example, recommended message sign character heights based on this guideline, according to viewing distance and number of display colours.

Number of Colours	Recommended Character Height in metres at 233m	300m
One	1.02	1.31
Three	1.82	2.35

Although both guidelines are for short viewing distances, both suggest that the present message signs have character sizes that are about half the size required for reliable legibility.

The effect of contrast on the legibility of luminous characters is equivalent to the luminance of the light sources used to form the characters. In the field trial, there were no luminance data available on the luminance of the different displays. Observers and research personnel alike noticed significant apparent differences in the brightnesses of the different displays. It was also noticed coincidentally that some sign brightnesses seemed quite dependent on viewing angle; that is, small changes in viewing position seemed to produce significant changes in the apparent brightness of some displays. In any case, observers often said that they could identify some displays more accurately than others because these displays were brighter than others. Indeed, if the different displays had had luminance strengths different from the ones they happened to have for the field trial, it is quite possible that an entirely different set of results would have been obtained. For this reason, it would have been preferable to have been able to control the luminance values of the different displays so that the results would be independent of display luminance.

Although the field trial provided information about the relative merits of the signs, as well as some useful information about general design features that affect the legibility of such signs, the field trial environment also put significant constraints on the testing methodology, as well as on the reliability of the testing procedures. The field environment did not permit much flexibility in setting up controlled viewing conditions, or in systematically varying critical variables such as sign luminance, viewing distance, and so on. The outdoor environment also places limitations on the type of apparatus that can be used, and on the precision with which measurements can be made. Finally, for observers, it is difficult to maintain constant viewing conditions, let alone constant standards of comfort.

Because of the constraints of the field environment, some of the inferences and conclusions drawn from the results of the trial represent products of circumstances that happened to be available in the environment of the trial rather than representing the products of a systematically controlled

experiment. Thus conclusions about such effects as font style should be treated as provisional conclusions until they can be confirmed by more systematic testing, such as could be done in a laboratory environment.

6.1 Conclusions

The following summarizes our conclusions from the trial.

1) The Dambach sign gave the optimum legibility performance for the symbols signs.

2) The Yvan Paque sign gave the optimum legibility performance for the numbers signs. The Dambach sign would also be a possible choice for the number displays provided that it could be shown that the 60 and 70 displays which were not provided by Dambach for the trial yielded legibility performance equal to that of the other Dambach displays.

3) For a choice of one manufacturer to provide both the symbol and the numeric displays, there were two candidates neither of which would be a clear choice on the basis of this trial. Yvan Paque would be one choice provided that Yvan Paque could implement the solid geometric form type of symbol display. Dambach would be the other choice provided that Dambach could implement satisfactory 60 and 70 sign displays.

4) The trial results suggested that the best font style for the number signs is a line stroke style constructed from closely spaced single lines of light elements. For the symbol signs, dense arrays of light elements forming solid geometric shapes appear to provide the best legibility.

5) The trial showed that the flashing displays significantly decrease legibility; and that the flashing mode does not convey any additional information not available in the steady mode.

6) Although it was not possible to systematically measure the effects of sign colour in the trial, it was evident that colour did not convey any extra

information; and to the degree that other research can be generalized to the message signs, colour probably decreased the legibilities of the signs.

7) Observers interpretations of the meanings of different signs were not consistent; and were often quite incorrect. This suggests that drivers would have to be educated about the meanings of the signs, and educated about driving behaviors that would be appropriate in response to the sign messages.

8) Compared with standards for constructing legible printed signs for use in other environments, the signs tested in this trial use character sizes that are about half the size that would be required for reliable legibility performance. In fact, at 300 metres, the number signs are equivalent to reading relatively small print on a printed page. Considering the environment in which the signs are to be used, there is little reason to be optimistic about their legibility characteristics. The obvious solution to this problem would be to increase the size of the signs, and to make them as bright as possible. If there are limits to the size and brightness of the signs, careful attention should be given to the font designs used for the signs.

9) The designs of the signs should be tested in a laboratory setting where the several variables that affect legibility can be properly controlled and manipulated. Laboratory testing would be critical to optimising such things as font design, colour, and size-luminance relationships.

7.0 REFERENCES

Kantowitz, Barry H. & Sorkin, Robert D. **Human Factors: Understanding People-System Relationships**, Wiley & Sons, New York, 1983.

Meister, D. **Human Engineering Data Base For Design And Selection of Cathode Ray Tube and Other Display Systems,** Elsevier, 1983.

Peters, G.A. & Adams, B.B. Three Criteria for readable panel markings cited in Salvendy, Gavriel (Ed) **Handbook of Human Factors**, Wiley & Sons, New York, 1987. pg 510.